

R and RStudio

(AST230) R for Data Science
Md Rasel Biswas



Introduction to R

- R is an extremely powerful programming language for statistical computing and graphics generation
- In 1993, **R** was created at the University of Auckland, New Zealand by two statistics professors
 - **Ross Ihaka** (University of Auckland, New Zealand)
 - **Robert Gentleman** (University of Waterloo, Canada)
- R is a dialect of *S Programming Language*



Introduction to R

- R is flexible and free to download (under GNU general public license), and has been widely used in academic environments over last two decades
- R is open source and is supported by an extensive user community
- R is currently maintained by the *R Development Core Team*
- *CRAN* (the Comprehensive R Archive Network) is a repository of additional R packages, contributed by the R user community



Why should we learn R

- R is open source and freely available.
- R is available for Windows, Mac and Linux operating systems.
- R has an extensive and highly flexible graphical facility capable of producing publication quality figures
- R has an expanding set of freely available 'packages' to extend R's capabilities.



Why should we learn R

Scripts

- R is a command line driven program. With R, all the steps used in the analysis (e.g. from reading the data to produce final results) can be saved and can redo the analysis without much effort
 - E.g. only running the scripts is needed, no need to remember the sequence of clicks used for the analysis
- Working with scripts makes the steps used in the analysis clear, and the code can be inspected by others (this will improve the codes and remove mistakes, if there is any!)
- Working with scripts helps to understand the associated statistical methods more clearly



Why should we learn R

Reproducibility

- The term reproducibility is used when someone else (including your future self!) can obtain the same results from the same data set when using the same analysis (coded in scripts)
- Now-a-days funding agency and peer-reviewed journals expect the analyses to be reproducible (journals often ask for the data and codes before publishing the accepted manuscripts)
- R becomes an integral part of reproducible research and it can be used to generate (dynamic) documents (e.g. manuscripts, report, etc.) from the codes (i.e. a small change of data, analysis, and organization can be updated automatically by running the scripts again)



Download R

- Download R installer from Comprehensive R Archive Networks (CRAN)
<https://cran.r-project.org>
- To download R installer for Windows OS
 - Click on ***Download R for Windows***
 - Click on ***Install R for the first time***

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.



R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to install R for the first time .
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.



R-4.1.0 for Windows (32/64 bit)

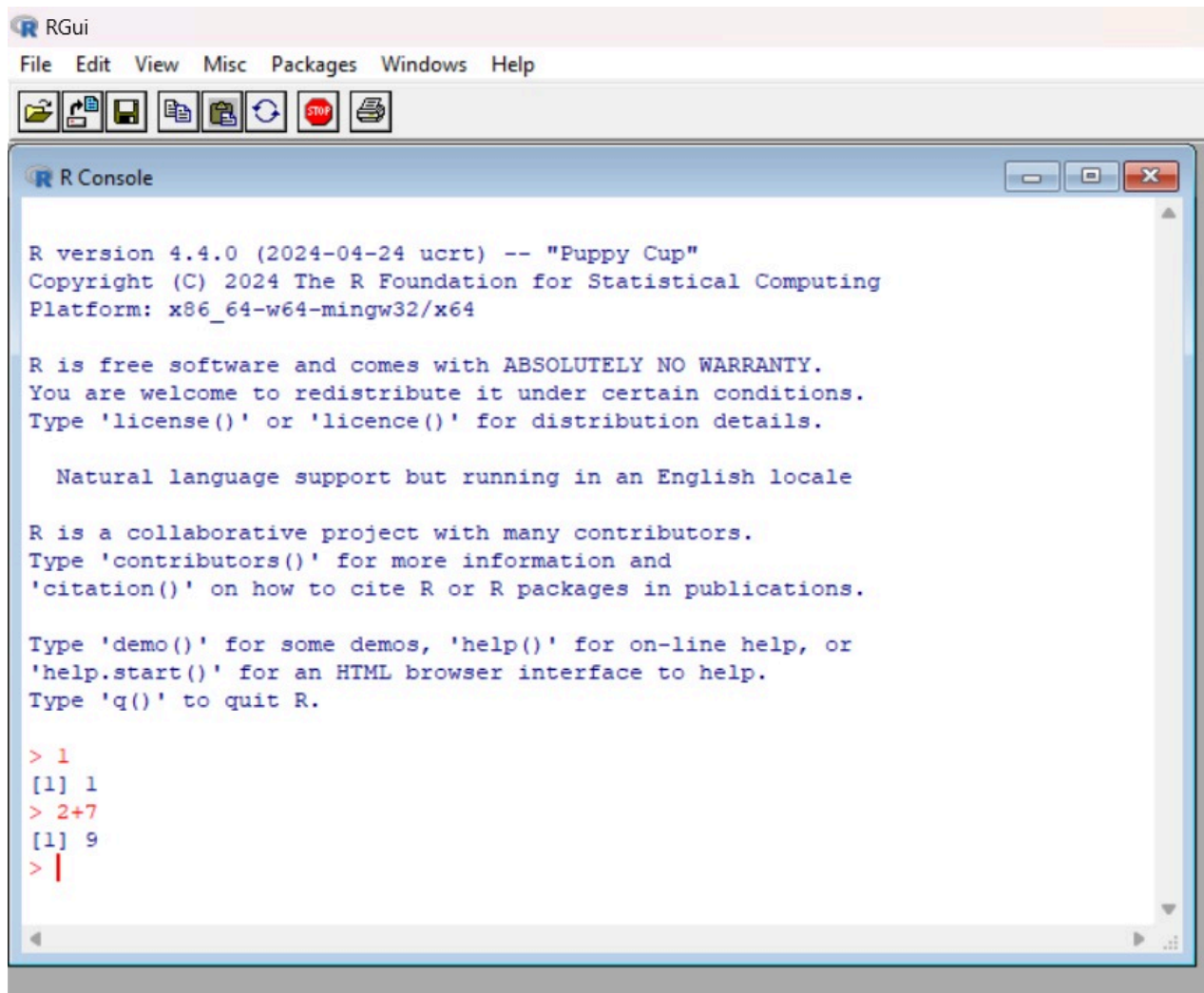
[Download R 4.1.0 for Windows](#) (86 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)



R interface



The screenshot shows the RGui application window. The title bar reads "RGui". The menu bar includes "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations (open, save, print, etc.). The main window is titled "R Console" and contains the following text:

```
R version 4.4.0 (2024-04-24 ucrt) -- "Puppy Cup"  
Copyright (C) 2024 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> 1  
[1] 1  
> 2+7  
[1] 9  
> |
```



Download RStudio

- Go to the page <https://posit.co/download/rstudio-desktop/> to download RStudio
- RStudio is an *integrated development environment (IDE)* for R. IDE is a GUI, where you can write your codes, see the results and also see the variables that are generated during the course of programming.

R is the language

RStudio is a software created to facilitate our use of R



RStudio interface

The screenshot displays the RStudio interface with four main panes:

- Source:** Shows the R script code for creating a ggplot2 plot. The code is:


```
1 library(ggplot2)
2 mpg_plot <- ggplot(mpg, aes(x = displ, y = hwy)) +
3   geom_point(aes(colour = class))
4
5 mpg_plot
6
```
- Console:** Shows the execution of the code in the R terminal. The output is:


```
> library(ggplot2)
> mpg_plot <- ggplot(mpg, aes(x = displ, y = hwy)) +
+   geom_point(aes(colour = class))
>
> mpg_plot
>
```
- Environments:** Shows the Global Environment with a table of objects:

Name	Type	Len...	Size	Value
mpg_plot	gg	9	29.1...	List of 9
- Output:** Displays a scatter plot of highway mileage (hwy) versus engine displacement (displ), colored by car class. The legend indicates the following classes: 2seater (red), compact (yellow), midsize (green), minivan (teal), pickup (cyan), subcompact (purple), and suv (pink).



RStudio interface

The RStudio user interface has 4 primary panes:

- **Source pane:** used to write and edit R codes and other related documents
- **Console pane:** This is the workhorse of R. This is where R evaluates all the code you write.
- **Environment pane,** containing the Environment, History, Connections, Build, VCS , and Tutorial tabs
- **Output pane,** containing the Files, Plots, Packages, Help, Viewer, and Presentation tabs



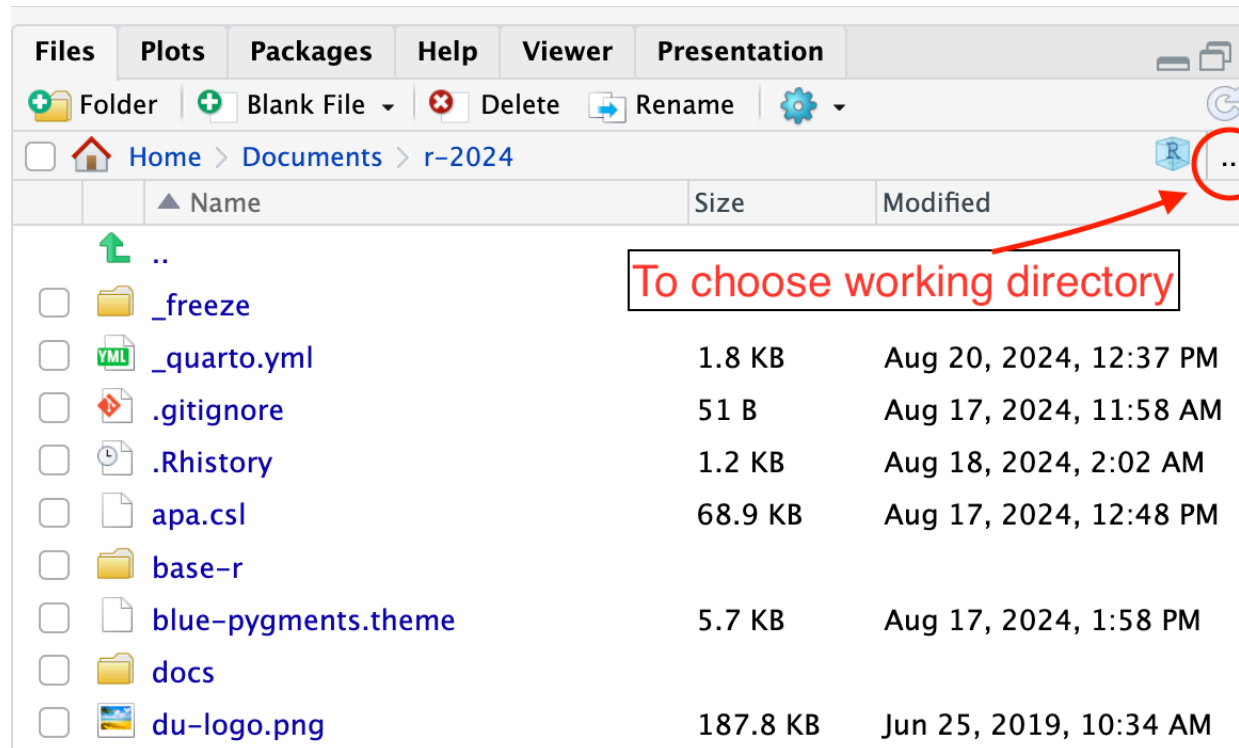
Set the working directory in RStudio

- R is always pointed at a directory on our computer. You can check the file path of your working directory by looking at bar at the top of the Console pane.
- We can also find out the working directory by running the `getwd()` function in **console**.
- We can set the working directory manually in two ways:
 1. The first way is to use the **console** and using the command `setwd("directory/path")`.
 - You can use this function `setwd()` and give the path of the directory which you want to be the working directory for RStudio, in the double quotes.



Set the working directory in RStudio

- The second way is to set the working directory from the GUI. Click on this 3 dots button which opens up a file browser, which will help you to choose your working directory.



Set the working directory in RStudio

Once you choose your working directory, you need to use this setting button in the more tab and click it and then you get a popup menu, where you need to select "*Set as working directory*"

